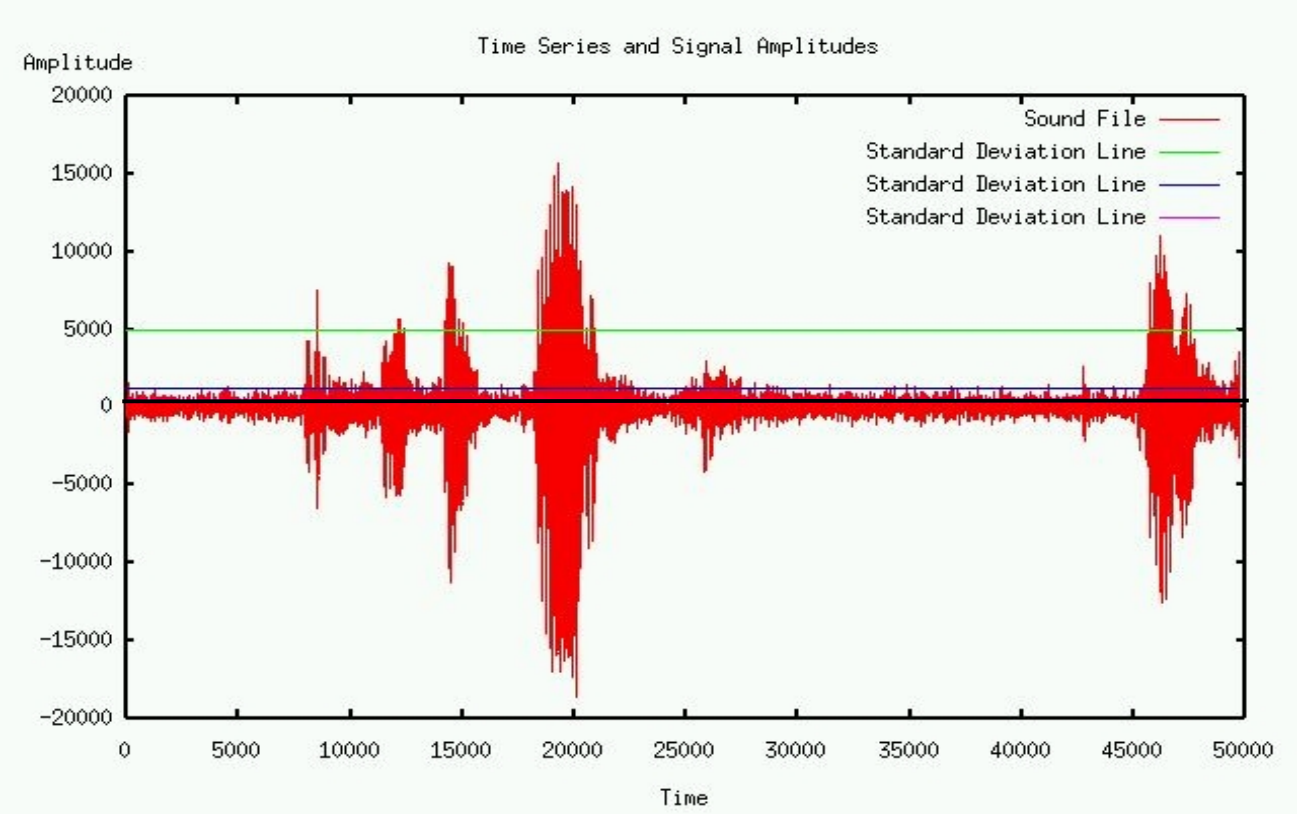
Speech Quality Metrics for Predicting Recognition Performance

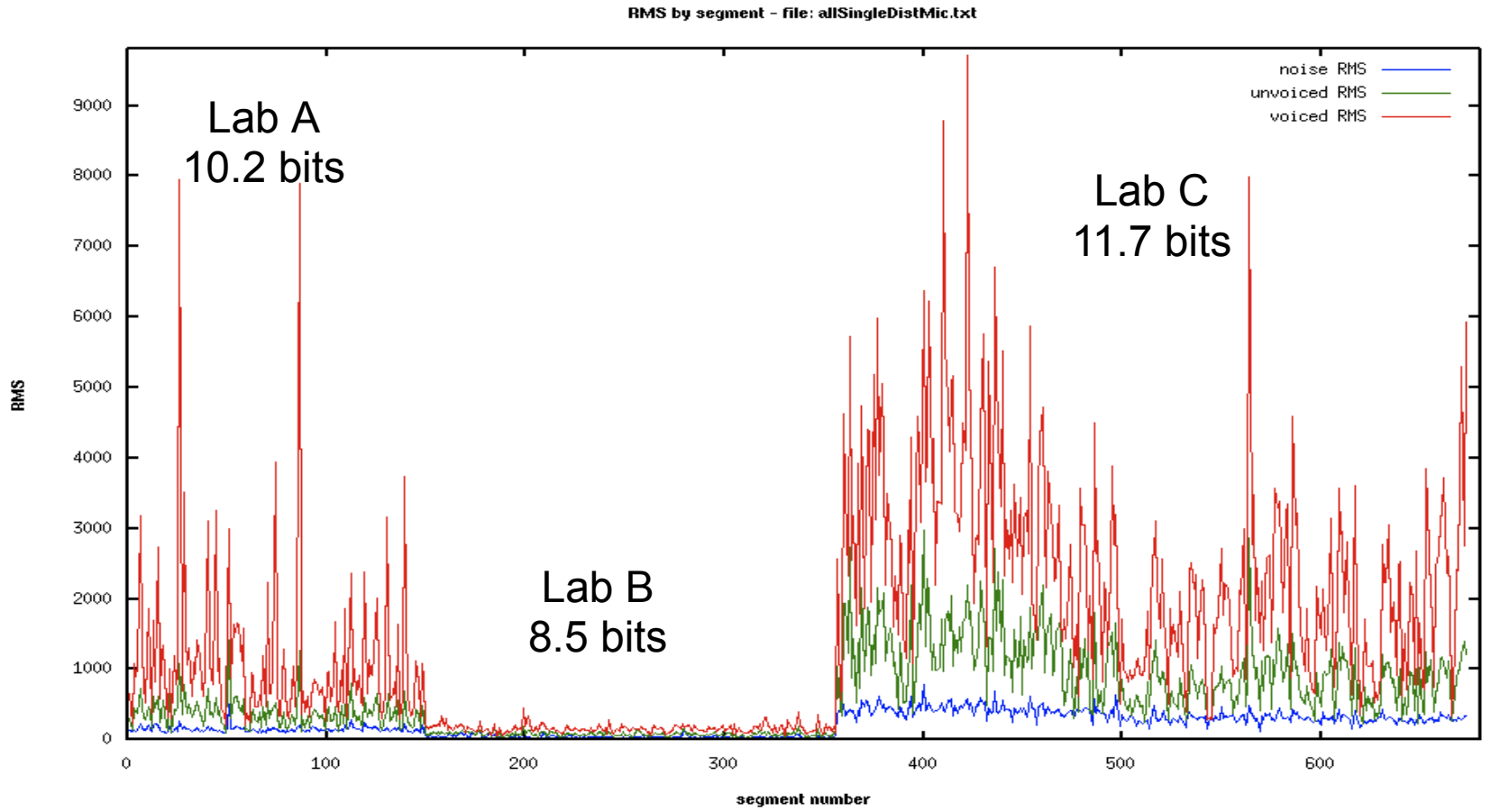
A Feasibility Study

Mathieu Hoarau and Vincent Stanford
Rich Transcription Evaluation, May 28, 2009
NIST ITL Information Access Division



Data Heterogeneity in a Previous RT Evaluation

Consistency and quality varied across labs



Background noise, unvoiced, and voiced speech (single distant microphone)

Speech Quality Metrics

- SNR Based:
  - Multi-band SNR metrics designed for clinical hearing assessment:
    - Articulation Index (AI) - French & Steinberg 1947
    - Speech Intelligibility Index (SII) - ANSI S3.5-1997 for stationary background noise
  - Broadband SNR metrics:
    - Bi-Gaussian – common mean, noise, and speech
    - Tri-Gaussian – common mean
- Signal information content:
  - Shannon Entropy

$$p(x|\sigma_n,\sigma_s)=\frac{P_n}{\sigma_n\cdot\sqrt{2\cdot\pi}}\exp^{\frac{(x/\sigma_n)^2}{2}}+\frac{P_s}{\sigma_s\cdot\sqrt{2\cdot\pi}}\exp^{\frac{(x/\sigma_s)^2}{2}}$$

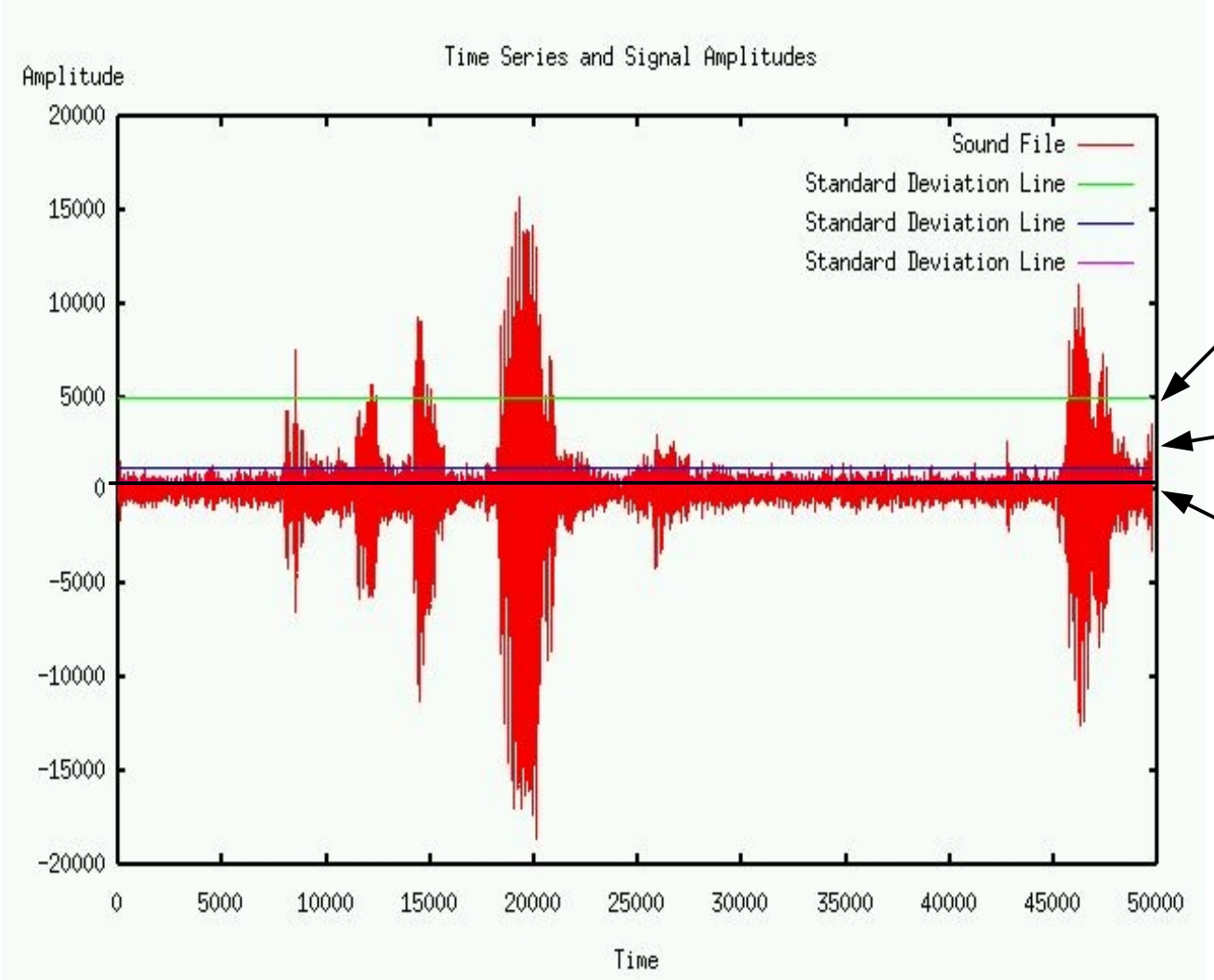
$$p(x|\sigma_n,\sigma_u,\sigma_v)=\frac{P_n}{\sigma_n\cdot\sqrt{2\cdot\pi}}\exp^{\frac{(x/\sigma_n)^2}{2}}+\frac{P_u}{\sigma_u\cdot\sqrt{2\cdot\pi}}\exp^{\frac{(x/\sigma_u)^2}{2}}+\frac{P_v}{\sigma_v\cdot\sqrt{2\cdot\pi}}\exp^{\frac{(x/\sigma_v)^2}{2}}$$

$$E=\sum_{i=-32768}^{32767} p_i\cdot\ln(p_i)$$

Data and Statistical Methods

- **Data sets subsets of:**
  - **RT-07 ASR system performance - single speaker utterances**
  - **2008 Follow-Up Speaker Recognition Evaluation – matched speakers for test probe and training**
- **Analysis Method:**
  - **Multiple Logistic Regression of recognition outcomes upon the speech quality metrics:**
    - **ASR - word level correct/incorrect**
    - **SID - correct identification of speaker**
- **Factor Analysis to test independence of the metrics set**

tri-Gaussian SNR Example



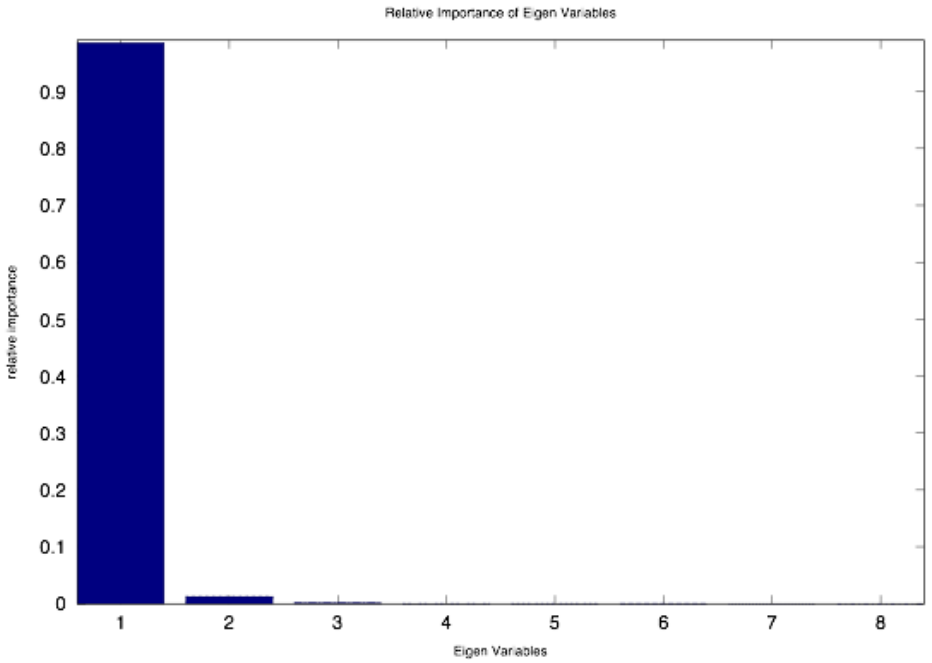
Voiced RMS

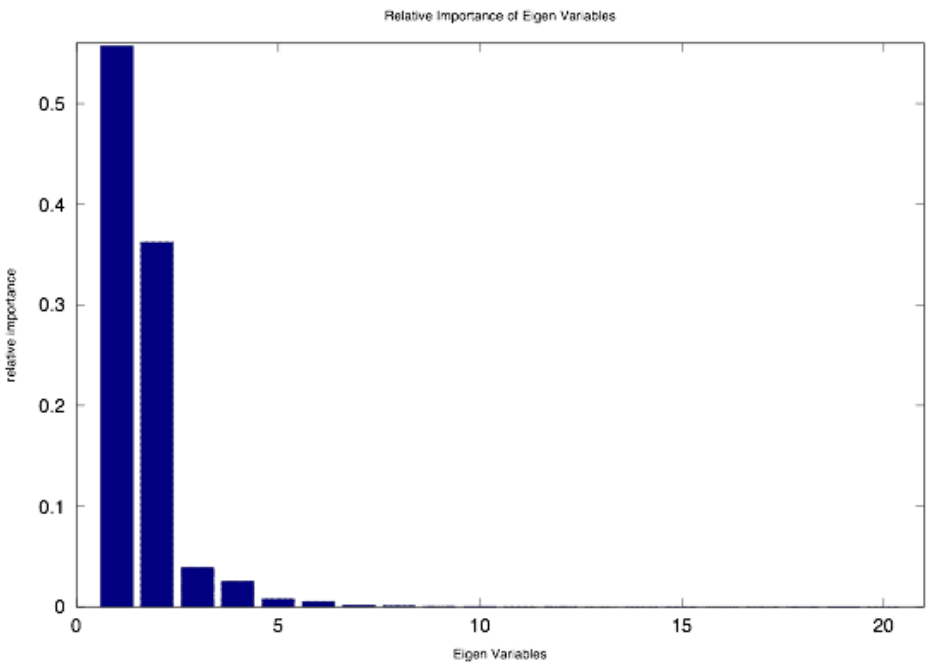
Unvoiced RMS

Background RMS

Factor Analysis of Results

Speech and Speaker Recognition Evaluation



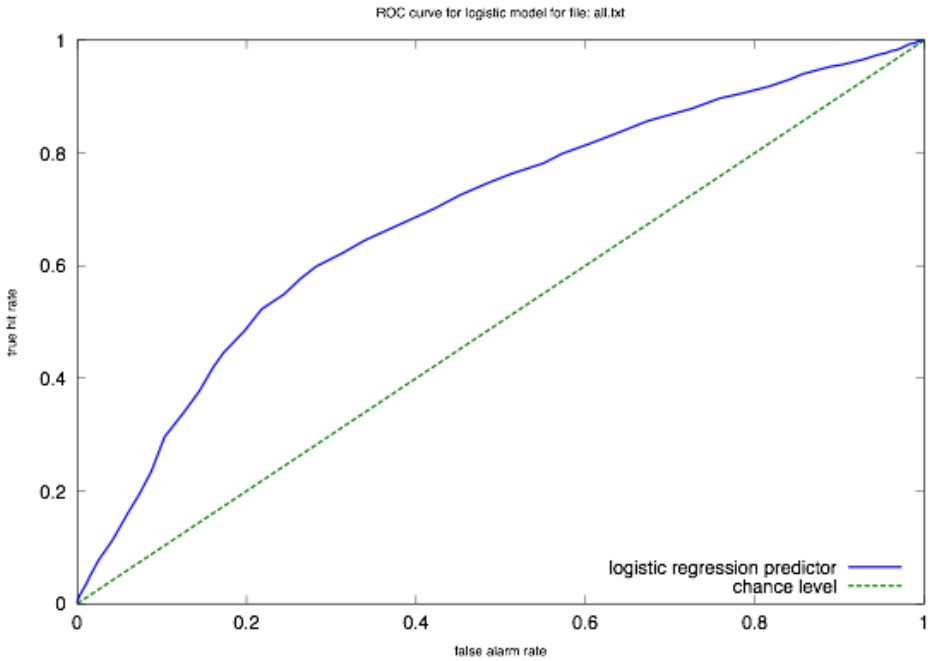


$$Cov(X) = \lambda_1 E_1^T \cdot E_1 + \ldots + \lambda_n E_n^T \cdot E_n$$

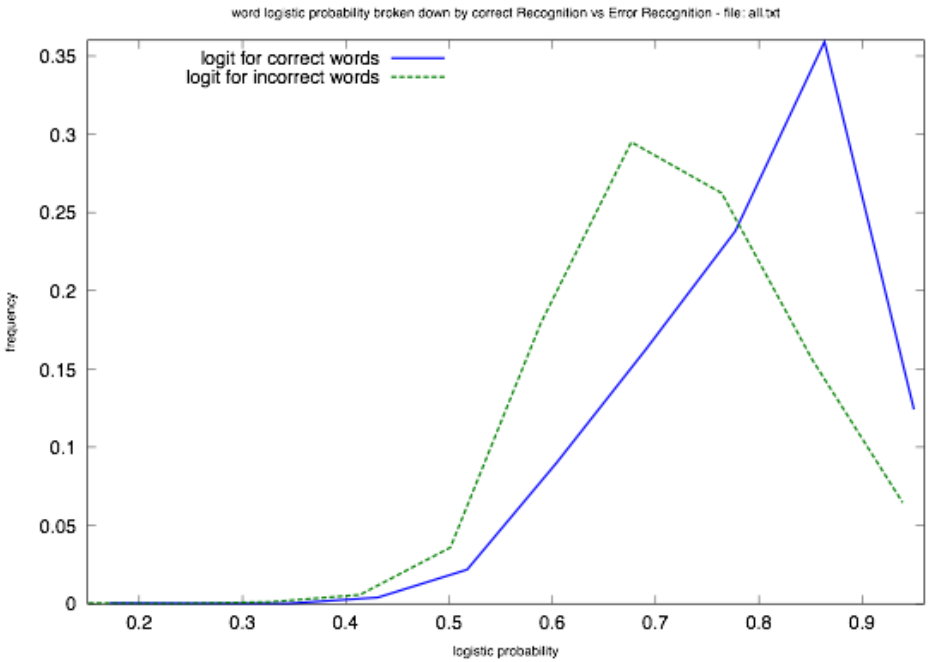
Conclusion: metric set well represented by one underlying factor
seen once in ASR test speech and once each in train and test pairs for SID

Speech Recognition Results

RT-07 Speech Recognition Evaluation



ROC Curve (diagonal=random)

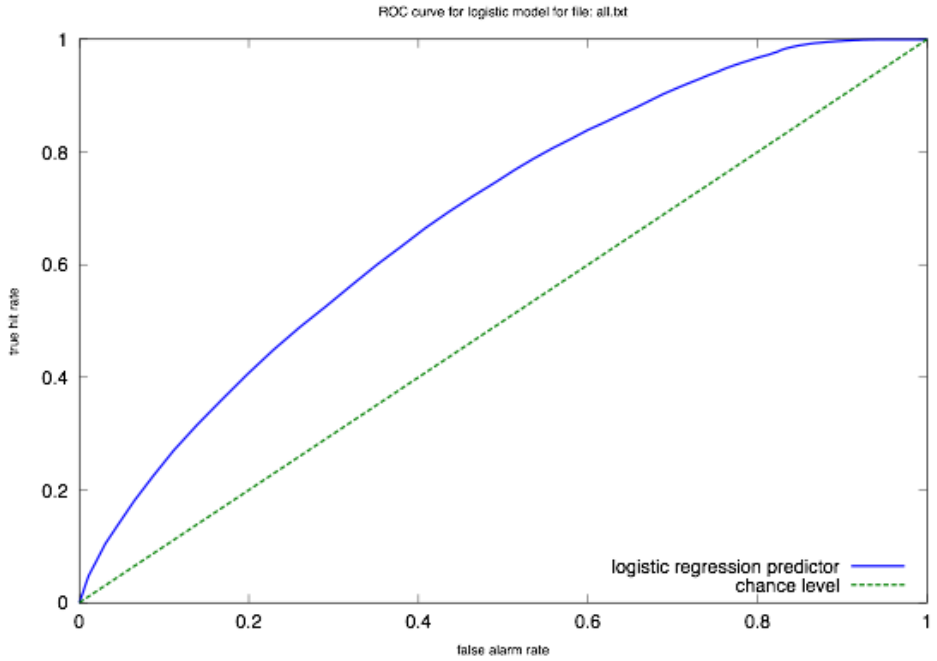


Distribution of Logistic Probabilities for correct vs incorrect ID

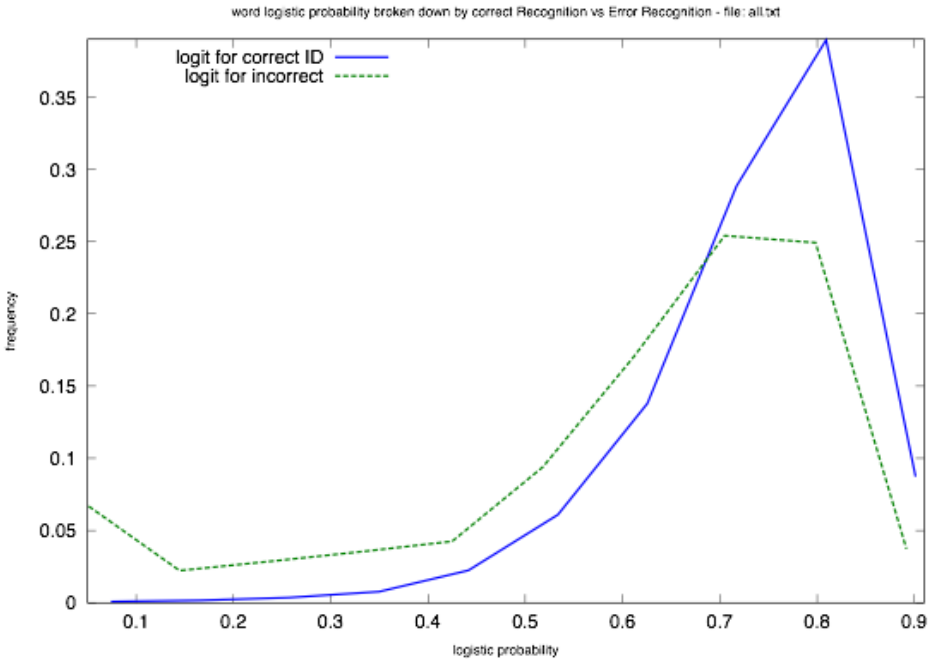
Head and Distant Microphones - Combined Metrics

Speaker Identification Results

2008 Follow-up Speaker Recognition Evaluation



ROC Curve (diagonal=random)



Distribution of Logistic Probabilities
for correct vs incorrect ID

Preliminary Conclusions

•It is feasible to predict recognition performance of complex biometric algorithms using simpler speech quality measurements.

•The metrics used in the present feasibility study:

- Are only moderately good (bad?) predictors of recognition performance
- SII and Multi-Gaussian SNR metrics capture substantially redundant information

•The present feasibility study did not include well-known features, e.g. Mel Frequency Cepstra used by the biometric algorithms.

•Speech quality does matter to recognition performance.

Questions for Further Research

- Can additional metrics that provide independent information and improved performance prediction be found?
- Is a quality measure evaluation track of interest?